

#2
DM

Attorney Docket No. 826.1763

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:

Isao NAMBA

Application No.: Unassigned

Group Art Unit: Unassigned

Filed: October 8, 2001

Examiner:

11046 U.S. PTO
09/971959
10/09/01

For: RETRIEVAL APPARATUS, RETRIEVAL METHOD AND RETRIEVAL PROGRAM

**SUBMISSION OF CERTIFIED COPY OF PRIOR FOREIGN
APPLICATION IN ACCORDANCE
WITH THE REQUIREMENTS OF 37 C.F.R. § 1.55**

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

In accordance with the provisions of 37 C.F.R. § 1.55, the applicant(s) submit(s)
herewith a certified copy of the following foreign application:

Japanese Patent Application No. 2001-108307

Filed: April 6, 2001

It is respectfully requested that the applicant(s) be given the benefit of the foreign filing
date(s) as evidenced by the certified papers attached hereto, in accordance with the
requirements of 35 U.S.C. § 119.

Respectfully submitted,

STAAS & HALSEY LLP

Date: October 8, 2001

By: _____

James D. Halsey, Jr.
Registration No. 22,729

700 11th Street, N.W., Ste. 500
Washington, D.C. 20001
(202) 434-1500

日 本 国 特 許 庁
JAPAN PATENT OFFICE

J1046 U.S. PTO

09/971959



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日
Date of Application:

2001年 4月 6日

出 願 番 号
Application Number:

特願2001-108307

出 願 人
Applicant(s):

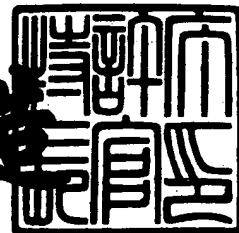
富士通株式会社

CERTIFIED COPY OF
PRIORITY DOCUMENT

2001年 7月 6日

特許庁長官
Commissioner,
Japan Patent Office

及川耕造



【書類名】 特許願

【整理番号】 0052787

【提出日】 平成13年 4月 6日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/30

【発明の名称】 検索装置、検索方法および検索プログラム

【請求項の数】 9

【発明者】

 【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

 【氏名】 難波 功

【特許出願人】

 【識別番号】 000005223

 【氏名又は名称】 富士通株式会社

【代理人】

 【識別番号】 100074099

 【住所又は居所】 東京都千代田区二番町8番地20 二番町ビル3F

 【弁理士】

 【氏名又は名称】 大菅 義之

 【電話番号】 03-3238-0031

【選任した代理人】

 【識別番号】 100067987

 【住所又は居所】 神奈川県横浜市鶴見区北寺尾7-25-28-503

 【弁理士】

 【氏名又は名称】 久木元 彰

 【電話番号】 045-573-3683

【手数料の表示】

 【予納台帳番号】 012542

 【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9705047

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 検索装置、検索方法および検索プログラム

【特許請求の範囲】

【請求項 1】 検索のための質問文を入力する質問文入力手段と、
検索対象のデータが格納されたデータベースを検索し、前記質問文入力手段によって入力された質問文に類似するデータを抽出する検索実行手段と、
前記検索実行手段によって抽出された検索結果に対して、前記検索実行手段による抽出に貢献した単語に関する貢献の度合を演算する単語貢献度演算手段と、
前記単語貢献度演算手段によって演算された貢献度を対応する単語とともに出力する単語貢献度出力手段とを備えたことを特徴とする検索装置。

【請求項 2】 前記単語貢献度出力手段は、前記貢献度に対応した表示形態で前記対応する単語を出力することを特徴とする請求項 1 に記載の検索装置。

【請求項 3】 前記検索実行手段は、前記質問文入力手段によって入力された質問文を単語に分割し、前記分割した単語毎に前記データベースを検索して類似するデータを抽出することを特徴とする請求項 1 乃至 2 の何れか 1 項に記載の検索装置。

【請求項 4】 前記単語貢献度演算手段は、
前記検索実行手段によって抽出された検索結果のうち、前記類似の度合が上位である集団と前記類似の度合が下位である集団とを取得する上位集団下位集団取得手段と、

前記分割した単語の前記上位集団に含まれる割合と前記下位集団に含まれる割合との差を求め、前記差を前記分割した単語の貢献度とする貢献度計算手段と

を備えたことを特徴とする請求項 3 に記載の検索装置。

【請求項 5】 検索により質問単語に類似するデータを抽出し、前記抽出された検索結果に対して、前記抽出に貢献した単語に関する貢献の度合（貢献度）を対応する単語とともに画面上に表示する検索装置において、

前記画面上に表示された単語を指定する単語指定手段と、

前記単語指定手段によって指定された単語に対して、前記検索のための重みを

付与する重み付け手段とを備えたことを特徴とする検索装置。

【請求項 6】 検索のための質問文を入力し、
検索対象のデータが格納されたデータベースを検索し、
前記入力された質問文に類似するデータを抽出し、
前記抽出された検索結果に対して、前記抽出に貢献した単語に関する貢献の度合を演算し、
前記演算された貢献度を対応する単語とともに出力することを特徴とする検索方法。

【請求項 7】 検索のための質問文を入力し、
前記入力された質問文を単語に分割し、
前記分割した単語毎に検索対象のデータが格納されたデータベースを検索し、
前記入力された質問文に類似するデータを抽出し、
前記抽出された検索結果に対して、前記抽出に貢献した単語に関する貢献の度合を演算し、
前記演算された貢献度を対応する単語とともに画面上に表示し、
前記画面上に表示された単語を指定し、
前記指定された単語に対して、前記検索のための重みを付与することを特徴とする検索方法。

【請求項 8】 コンピュータに、
検索のための質問文を入力する機能、
検索対象のデータが格納されたデータベースを検索する機能、
前記入力された質問文に類似するデータを抽出する機能、
前記抽出された検索結果に対して、前記抽出に貢献した単語に関する貢献の度合を演算する機能、
前記演算された貢献度を対応する単語とともに出力する機能、
を実現させるための検索プログラム。

【請求項 9】 検索により質問単語に類似するデータを抽出し、前記抽出された検索結果に対して、前記抽出に貢献した単語に関する貢献の度合（貢献度）を対応する単語とともに画面上に表示するコンピュータに、

前記画面上に表示された単語を指定する機能、

前記指定された単語に対して、前記検索のための重みを付与する機能、
を実現させるための検索プログラム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、自然言語で記述された検索要求に対して検索を行ったり、あるいは入力された単語列からなる質問に対して検索を行ったりした検索結果を、文書類の類似度順に順位をつけてユーザーに提示する類似文書検索装置に関する。

【0002】

【従来の技術】

従来、類似文書検索装置は、ユーザーが入力した検索要求に対して最も類似度が高いと判断される文書から順に順位を付け、その検索結果を提示している。

【0003】

類似文書検索に関する技術は、例えば、Managing Gigabyte s, Compressing and Indexing Documents and Images, 141ページ乃至148ページに開示されている。

【0004】

文書の類似度は、一般には検索要求単語の検索対象文書中での重要度 (i d f) とそれぞれの文書中での重要度 (t f) をかけることにより求められる。

【0005】

【発明が解決しようとする課題】

しかしながら、通常ユーザーには、検索結果中での単語の重要度は明らかにされることはないし、また類似度検索式はユーザーには公開もされないし、その作用がユーザーには理解し難いものである。

【0006】

そのため、検索要求に対して予想と異なった文書が上位に上がってきた場合には、ユーザーは自分で類似文書検索装置の動きを検索結果から予測して単語を置き換えて検索するか、質問文（検索要求）を作成し直して再度検索する以外に方

法はないという問題があった。

【0007】

本発明は、上記問題点に鑑みてなされたもので、どのような単語が検索で重視されたかをユーザーに通知するができ、意図と違った形の類似検索の結果が返ってきてても、単語に対する重みを容易に修正することが可能となる検索装置、検索方法および検索プログラムを提供することを目的とする。

【0008】

【課題を解決するための手段】

従来の類似文書検索装置で上述のような問題があるのは、なぜそのような結果が返ってくるのかユーザーには直感的に理解し難いからである。これは検索がうまくいっている場合には問題はない。しかし、類似文書検索装置がユーザーの思いもよらない結果を返した場合にはユーザーとしてはどのような対応をとればよいか明確ではない。

【0009】

以下の文書は、TREC (Text Retrieval Conference) からの質問文である。

Identify documents that discuss the current status of hybrid automobile engines, (i.e., cars fueled by something other than gasoline only). A relevant document may include research on non-gasoline powered engines or prototypes that may be fueled by natural gas, methanol, alcohol; cost to the consumer; health benefits derived; and shortcomings in horsepower and passenger comfort.

この質問文は整理された形でサーチャーに届くような検索要求の形をとっているが、この文書をそのまま検索要求として、自然言語を入力とする類似文書検索装置に投入すると、「fuel」か「car」を主題とした文書が上位に並び、「hybrid fueled car」に関する結果は上位には並ばない。この場合に検索結果を良く見れば、「hybrid」という単語の類似文書検索装置内部での重要度が低いことと、「car」、「fuel」という単語があるとそれらが中心となった文書が上位にくるために、望んだ結果が得られないことがわかる。しかしながらこのようなことを一般ユーザーが判断することは容易ではない。

【0010】

また、このような場合に、「hybrid」にユーザー側で重みをおけば適切な検索結果を得られる。しかし、細かい数値で単語の重みを指定するにせよ、現在重みが置かれている単語「fuel」、「car」とバランスした重みにするための計算方法、指定方法が存在しないため、容易に重み付けを行なうことができない。

【0011】

本発明は、上記課題を解決するため、下記のような構成を採用した。

(1) 検索結果全体の傾向を表わす形で類似文書検索装置側が判定した単語の効き（貢献度）を階調表示することにより、どのような単語が検索で重視されたかをユーザーに通知する。

(2) 指定された単語の重み付けを変更し、検索結果に対して適切な貢献度が得られるように修正するインターフェイス並びに内部的な計算機能を実現する。これによりユーザーは、意図と違った形の類似検索の結果が返ってきても、単語に対する重みを容易に修正することが可能となる。

【0012】

すなわち、本発明の一態様によれば、本発明の検索装置は、検索のための質問文を入力する質問文入力手段と、検索対象のデータが格納されたデータベースを検索し、上記質問文入力手段によって入力された質問文に類似するデータを抽出する検索実行手段と、上記検索実行手段によって抽出された検索結果に対して、上記検索実行手段による抽出に貢献した単語に関する貢献の度合を演算する単語貢献度演算手段と、上記単語貢献度演算手段によって演算された貢献度を対応する単語とともに出力する単語貢献度出力手段とを備えたことを特徴とする。

【0013】

また、本発明の検索装置は、上記単語貢献度出力手段が、上記貢献度に対応した表示形態で上記対応する単語を出力することが望ましい。

また、本発明の検索装置は、上記検索実行手段が、上記質問文入力手段によって入力された質問文を単語に分割し、上記分割した単語毎に上記データベースを検索して類似するデータを抽出することが望ましい。

【0014】

また、本発明の検索装置は、上記単語貢献度演算手段が、上記検索実行手段によって抽出された検索結果のうち、上記類似の度合が上位である集団と上記類似の度合が下位である集団とを取得する上位集団下位集団取得手段と、上記分割した単語の上記上位集団に含まれる割合と上記下位集団に含まれる割合との差を求め、上記差を上記分割した単語の貢献度とする貢献度計算手段とを備えたことが望ましい。

【0015】

また、本発明の一態様によれば、本発明の検索装置は、検索により質問単語に類似するデータを抽出し、上記抽出された検索結果に対して、上記抽出に貢献した単語に関する貢献の度合（貢献度）を対応する単語とともに画面上に表示する検索装置であって、上記画面上に表示された単語を指定する単語指定手段と、上記単語指定手段によって指定された単語に対して、上記検索のための重みを付与する重み付け手段とを備えたことを特徴とする。

【0016】

【発明の実施の形態】

以下、本発明の実施の形態を、図面を参照しながら詳細に説明する。

図1は、本発明を適用した検索装置の機能構成図である。

【0017】

図1において、検索装置1は、質問文入力手段2、検索実行手段4、単語貢献度演算手段5、単語指定手段9、重み付け手段10、出力手段13を備え、検索により質問単語に類似するデータを抽出し、上記抽出された検索結果に対して、上記抽出に貢献した単語に関する貢献の度合（貢献度）を対応する単語とともに画面上に表示する。

【0018】

そして、検索実行手段4は、再検索実行手段11を備え、単語貢献度演算手段5は、上位集団下位集団取得手段7と貢献度計算手段8とを備え、出力手段13は、単語貢献度出力手段6と検索結果出力手段12とを備える。

【0019】

上記質問文入力手段2は、検索のための質問文を入力する。

上記検索実行手段4は、検索対象のデータが格納されたデータベース3を検索し、上記質問文入力手段2によって入力された質問文に類似するデータを抽出する。

【0020】

また、上記検索実行手段4は、上記質問文入力手段2によって入力された質問文を単語に分割し、上記分割した単語毎に上記データベース3を検索して類似するデータを抽出する。

【0021】

上記単語貢献度演算手段5は、上記検索実行手段4によって抽出された検索結果に対して、上記検索実行手段4による抽出に貢献した単語に関する貢献の度合を演算する。

【0022】

上記単語貢献度出力手段6は、上記単語貢献度演算手段5によって演算された貢献度を対応する単語とともに、CRTディスプレイ、液晶ディスプレイ、プラズマディスプレイ等の表示装置やレーザービームプリンタ、インクジェットプリンタ等の印刷装置等へ出力する。出力は、貢献度に対応した表示形態、貢献度に対応したフォント、貢献度に対応した文字サイズ、貢献度に対応した色等で行なう。

【0023】

上記上位集団下位集団取得手段7は、上記検索実行手段4によって抽出された検索結果のうち、上記類似の度合が上位である集団と上記類似の度合が下位である集団とを取得する。

【0024】

上記貢献度計算手段8は、上記分割した単語の上記上位集団に含まれる割合と上記下位集団に含まれる割合との差を求め、上記差を上記分割した単語の貢献度とする。

【0025】

上記単語指定手段9は、上記画面上に表示された単語を指定する。

重み付け手段10は、上記単語指定手段9によって指定された単語に対して、

上記検索のための重みを付与する。

【0026】

また、上記重み付け手段10は、上記検索結果のデータが最上位になるように重みを付与する。

また、上記重み付け手段10は、上記検索結果のうち所定のデータに対して、順に、上記検索結果のデータが最上位になるように重みを付与し、各データに対する重みの平均値を上記単語の重み値とする。

【0027】

上記再検索実行手段11は、検索対象のデータが格納されたデータベース3を検索し、上記重み付け手段10によって重みが付与された単語に類似するデータを抽出する。

【0028】

なお、検索手段1は、上記データベース3を備えるようにしてもよい。

図2は、本発明の原理説明図である。

図2において、質問文入力部21は、ユーザーの入力した質問文を受理する。

【0029】

検索実行部22は、ユーザーの入力した質問文に基づき検索を実行する。

単語貢献度計算部23は、検索結果に対して質問文中の単語(複数)の効き目を計算する。

【0030】

単語貢献度表示部24は、検索要求中の単語に対する貢献度を表示する。

単語重み修正要求部25は、ユーザーが単語の重みの修正要求を受け付ける。

修正単語重み計算部26は、ユーザーにより指定された単語の効きが高くなるように単語重みを計算し、単語に重みを付加する。

【0031】

図3は、検索結果中の単語の貢献度を表示するまでの処理の流れである。

ステップS1において、ユーザーからの検索要求を受理(入力)する。

ステップS2において、ステップS1入力された検索要求に対して、あるいは適宜、検索要求を単語に分割して、類似検索を実行し、関連度の高い文書順に結

果を得る。

【0032】

ステップS3において、ステップS2における検索による検索結果の上位N件の文書を取得する。これは最も検索要求に合致していると検索装置が判断した文書とみなせる。これらN件の文書の集団をAとする。ここで、上位N件とは、1位からN位までとは限らず、例えば、1位ずつスキップしたN件分であってもよい。

【0033】

ステップS4において、検索結果のM番目にランクされた文書よりP件の文書を取得する。これらは検索結果として引けてきたが、検索結果の中ではさほど重要度が高くなかった文書とみなせる。これらP件の文書の集団をBとする。ここで、P件とは、M位から(M+P-1)位までのP件とは限らず、例えば、1位ずつスキップしたP件分であってもよい。また、Mは、Nより大きくても小さくてもよい。

【0034】

ステップS5において、検索要求より1単語取り出す。

ステップS6において、処理対象の単語が最後の単語か否かを判定する。最後の単語であれば処理はステップS9に移る。それ以外の場合には、処理はステップS7に移る。

【0035】

ステップS7において、処理対象の単語がAとBとに含まれる割合をそれぞれ計算する。この貢献度の式としては、Term selection value(Okapi system), Term selection value(Boughamformula), Relevance feedbackformula(Rocchio)などを利用することが考えられる。

【0036】

なお、Term selection valueについては、NIST (National Institute of Standards and Technology) 発行の「The Eighth Text Retrieval Conference (TR EC-8)」の151ページ乃至161ページに詳細が記載されている。

【0037】

次に、ステップS8において、ステップS7の結果を貢献度とみなして単語に付与する。そののちステップS5に戻る。

上記ステップS5乃至ステップS8はループ処理であり、検索要求中からそれぞれの単語を取り出し、それらに対して、貢献度を計算する。

【0038】

そして、ステップS9において、単語の貢献度の最大値(Max)と最小値(Min)を求める。

ステップS10において、貢献度を階調表示するために、MinとMaxとの間を一定の階調に分ける。階調の設定は等分割などの方法が考えられる。

【0039】

ステップS11において、階調に合わせて表示のためのフォントを設定する。

ステップS12において、単語ごとにその貢献度に応じた対応したフォントで検索結果に対する単語の貢献度を表示する。

【0040】

図4は、検索結果中の単語の貢献度をユーザーが変更し、その変更内容を反映するように単語に重みを付ける処理の流れを示すフローチャートである。

ステップS11において、貢献度を変更する単語をユーザーが指定する。

【0041】

ステップS12において、類似検索を実行する。

ステップS13において、 $k=1$ とし、ステップS14において、 $L=n$ とする。ここで n は、上位 n 件分の n である。

【0042】

上記ステップS13およびステップS14は、 k が L となるまで、検索結果より k 番目の順位となった文書を取り出し、単語に補正の重みを付ける計算を行う。

【0043】

ステップS15において、現在の対象文書が最後であるか否か($K=L$ か否か)を判定する。

最後までいっているのであれば、処理は終了する。それ以外の場合には処理は

ステップS16に進む。

【0044】

ステップS16において、処理対象としてk番目の文書をとる。

ステップS17において、k番目の文書に貢献度を変更したい単語が含まれているかどうか確かめる。

【0045】

含まれていなければ処理はステップS22へ進み、それ以外の場合にはステップS18へ移る。

ステップS18において、現在k番目に位置する文書の順位が1位になるには、候補単語にどの程度重みをつければよいか計算する。

【0046】

ステップS19において、既に単語に対して重みが設定されているかどうかにより処理が分かれる。もし重みが設定されていないのであれば、重みにk番目の文書の処理結果を設定する。それ以外の場合には、既に処理した結果と平均をとる。

【0047】

ステップS10において、重みとして、k番目の文書に求めた単語の重みを設定する。

ステップS11において、重みとして、k番目の文書に求めた単語の重みと以前の重みの平均をとる。

【0048】

いずれの例でも修正して求めた重みには過度の強さを持たせないため、上位の文書から求めた重みほど利きが強くなるようにすることが必要である。下位の文書から上位の文書に上げるために求めた重みの影響が大きくなるようにするためには、求めた重みに $1 / \log$ （現在の処理対象文書の順位）などをかけて、補正することが考えられる。

【0049】

上記ステップS19乃至ステップS21は、平均を取る処理である。

ステップS22において、次の文書を処理するためにkをひとつ進める。その

のち処理はステップ S 1 6 に戻る。

【0050】

上記ステップ S 1 5 乃至ステップ S 2 2 は、ループであり、上位 n 件の文書を 1 件ずつとり、指定された単語の貢献度が上がるように単語に補整の重みをつける処理を行なう。

【0051】

以下の実施の形態においては、自然言語で記述された検索要求で英語文書に対して検索を行ない、検索結果に対する単語の貢献度をフォント表示し、ユーザーに対して検索装置の単語の重みを 4 階調のフォントの大きさとして示している。

【0052】

図 5 は、検索要求の例を示す図であり、検索したいものを規定した文書である。文書自体は TREC (Text REtrieval Conference) の質問文セットの 385 番より取っている。

【0053】

上記質問文は、「Hybrid car」に対する検索要求そのものと検索されるべき文書の詳細をあわせて記述したものである。

図 6 は、貢献度の計算式とパラメーターの設定値を示す図である。

【0054】

貢献度の計算として、関連度が高そうな文書の集合に上位 10 件 (図 3 中の A)、関連度が低そうな文書の集合 (図 3 中の B) に上位 800 件からの 200 文書を想定している。この場合には、検索結果として 1000 件以上が想定されているが、検索結果が 1000 件以下であれば、取得された結果で最後から数えて 2 割を B とするといった処理が考えられる。

【0055】

貢献度の計算式 (割合判定式) としては、Term Selection Value (Boughamformula) を用いているが、計算式には任意のものを用いることが可能である。

図 7 は、貢献度の計算結果を示す図である。

【0056】

図 7 において、「fuel」が貢献度 7.2 で最も高く、「hybrid」は検索要求中

での貢献度が低く（0.6）、意図した結果とずれた文書がとれてきたであろうことがわかる。

【0057】

図8は、貢献度を4階調のフォント表示とした図である。

最高点と最低点の間を等分に4つに分け、8ポイントづつフォントサイズを変えていき、単語を貢献度順にグループ化したものである。これにより検索結果では、「Fuel」、「methanol」、「cars」が支配的であることが見て取れる。

【0058】

図8に示した表示例では等分割したが、人目で理解できるように貢献度の高い3単語を最初のグループとし、それ以外を等分割するといった方法も考えられる。

【0059】

図9は、単語貢献度への変更要求の例を示す図である。

貢献度を変更して、望ましい結果を取ろうとするユーザーの試みをであり、入力インターフェイスに対する動作を表わしている。

【0060】

上位の単語の内、「methanol」は関係が薄そうなので削除し、「Fuel」の貢献度を下げる。またキーワードである「hybrid」の貢献度を1位になるように設定する。

【0061】

図10は、再計算対象となる類似検索で上位にきた文書中に各単語がどのように含まれ、どのような得点となっているかを示す図である。

図10において、文書に対する得点は各単語の得点を加算したものである。横の数字は各文書が何番目にランクされたかを示している。例えば、1番目にランクされた文書には、「Cars」の得点が「5」で、「hybrid」の得点が「0」等となっている。

【0062】

図11は、再計算の処理を示す図である。

現在8位（番目）にある「hybrid」を含む文書を1位（番目）に持ってくるに

は、「hybrid」の点に「9」をかければよいので、これを重みとする。ただし、すべての順位の文書にこの重みを直接用いると、総合点が低い文書でさえ、1単語の重みだけで順位を押し上げることになるため、順位が下がるにつれ重みの反映を弱くするためにここでは \log （文書の順位）をかけている。

【0063】

「fuel」の場合はこれを逆に行っており、1位の文書を10位に下げるための重みを逆に計算している。ここでは最低点を1としており、この値まで下げるために「fuel」にかける1以下の重みを算出している。

【0064】

図12は、重みの補正後の検索結果に対する貢献度を示す図である。

「hybrid」の得点が5.3となり、他の検索単語とバランスしながら貢献度で1位にきているのがわかる。

【0065】

図13は、重み変更後の表示結果を示す図である。

重みの補正により上位にきた「hybrid」の表示サイズが大きくなっているのが分かる。

【0066】

上述のように、本発明の実施の形態を、図面を参照しながら説明してきたが、本発明が適用される検索装置は、その機能が実行されるのであれば、上述の実施の形態に限定されることなく、単体の装置であっても、複数の装置からなるシステムあるいは統合装置であっても、LAN、WAN等のネットワークを介して処理が行なわれるシステムであってもよいことは言うまでもない。

【0067】

また、図14に示しように、バス140に接続されたCPU141、ROMやRAMのメモリ142、入力装置143、出力装置144、外部記録装置145、媒体駆動装置147、可搬記録媒体146、ネットワーク接続装置148で構成されるシステムでも実現できる。すなわち、前述してきた実施の形態のシステムを実現するソフトウェアのプログラムコードを記録したROMやRAMのメモリ142、外部記録装置145、可搬記録媒体146を、検索装置に供給し、その

検索装置のコンピュータがプログラムコードを読み出し実行することによっても、達成されることは言うまでもない。

【0068】

この場合、可搬記録媒体146等から読み出されたプログラムコード自体が本発明の新規な機能を実現することになり、そのプログラムコードを記録した可搬記録媒体146等は本発明を構成することになる。

【0069】

プログラムコードを供給するための可搬記録媒体146としては、例えば、フロッピーディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、DVD-ROM、DVD-RAM、磁気テープ、不揮発性のメモリーカード、ROMカード、電子メールやパソコン通信等のネットワーク接続装置148（言い換えれば、通信回線）を介して記録した種々の記録媒体などを用いることができる。

【0070】

また、図15に示すように、コンピュータ150がメモリ151上に読み出したプログラムコードを実行することによって、前述した実施の形態の機能が実現される他、そのプログラムコードの指示に基づき、コンピュータ150上で稼動しているOSなどが実際の処理の一部または全部を行ない、その処理によっても前述した実施の形態の機能が実現される。

【0071】

さらに、可搬型記録媒体152から読み出されたプログラムコードやプログラム（データ）提供者から提供されたプログラム（データ）153が、コンピュータ150に挿入された機能拡張ボードやコンピュータ150に接続された機能拡張ユニットに備わるメモリ151に書き込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行ない、その処理によっても前述した実施の形態の機能が実現され得る。

【0072】

すなわち、本発明は、以上に述べた実施の形態に限定されるものではなく、本

発明の要旨を逸脱しない範囲内で種々の構成または形状を取ることができる。

ここで、上述した実施の形態の特徴を列挙すると、以下の通りである。

(付記 1) 検索のための質問文を入力する質問文入力手段と、

検索対象のデータが格納されたデータベースを検索し、前記質問文入力手段によって入力された質問文に類似するデータを抽出する検索実行手段と、

前記検索実行手段によって抽出された検索結果に対して、前記検索実行手段による抽出に貢献した単語に関する貢献の度合を演算する単語貢献度演算手段と、

前記単語貢献度演算手段によって演算された貢献度を対応する単語とともに出力する単語貢献度出力手段とを備えたことを特徴とする検索装置。

(付記 2) 前記単語貢献度出力手段は、前記貢献度に対応した表示形態で前記対応する単語を出力することを特徴とする付記 1 に記載の検索装置。

(付記 3) 前記単語貢献度出力手段は、前記貢献度に対応したフォントで前記対応する単語を出力することを特徴とする付記 1 に記載の検索装置。

(付記 4) 前記単語貢献度出力手段は、前記貢献度に対応した文字サイズで前記対応する単語を出力することを特徴とする付記 1 または 3 に記載の検索装置。

(付記 5) 前記単語貢献度出力手段は、前記貢献度に対応した色で前記対応する単語を出力することを特徴とする付記 1、3 または 4 に記載の検索装置。

(付記 6) 前記単語貢献度出力手段は、表示装置であることを特徴とする付記 1 乃至 5 の何れか 1 項に記載の検索装置。

(付記 7) 前記検索実行手段は、前記質問文入力手段によって入力された質問文を単語に分割し、前記分割した単語毎に前記データベースを検索して類似するデータを抽出することを特徴とする付記 1 乃至 6 の何れか 1 項に記載の検索装置。

(付記 8) 前記単語貢献度演算手段は、

前記検索実行手段によって抽出された検索結果のうち、前記類似の度合が上位である集団と前記類似の度合が下位である集団とを取得する上位集団下位集団取得手段と、

前記分割した単語の前記上位集団に含まれる割合と前記下位集団に含まれる割合との差を求め、前記差を前記分割した単語の貢献度とする貢献度計算手段と

を備えたことを特徴とする付記 7 に記載の検索装置。

(付記 9) 検索により質問単語に類似するデータを抽出し、前記抽出された検索結果に対して、前記抽出に貢献した単語に関する貢献の度合(貢献度)を対応する単語とともに画面上に表示する検索装置において、

前記画面上に表示された単語を指定する単語指定手段と、

前記単語指定手段によって指定された単語に対して、前記検索のための重みを付与する重み付け手段とを備えたことを特徴とする検索装置。

(付記 10) 検索のための質問文を入力する質問文入力手段と、

前記質問文入力手段によって入力された質問文を単語に分割し、前記分割した単語毎に検索対象のデータが格納されたデータベースを検索し、前記質問文入力手段によって入力された質問文に類似するデータを抽出する検索実行手段と、

前記検索実行手段によって抽出された検索結果に対して、前記検索実行手段による抽出に貢献した単語に関する貢献の度合を演算する単語貢献度演算手段と、

前記単語貢献度演算手段によって演算された貢献度を対応する単語とともに画面上に表示する単語貢献度出力手段と、

前記画面上に表示された単語を指定する単語指定手段と、

前記単語指定手段によって指定された単語に対して、前記検索のための重みを付与する重み付け手段とを備えたことを特徴とする検索装置。

(付記 11) 前記重み付け手段は、前記検索結果のデータが最上位になるように重みを付与することを特徴とする付記 9 または 10 に記載の検索装置。

(付記 12) 検索対象のデータが格納されたデータベースを検索し、前記重み付け手段によって重みが付与された単語に類似するデータを抽出する再検索実行手段を備えたことを特徴とする付記 9 乃至 11 の何れか 1 項に記載の検索装置。

(付記 13) 検索対象のデータが格納されたデータベースを検索し、前記重み付け手段によって重みが付与された単語に類似するデータを抽出する再検索実行手段を備え、

前記重み付け手段は、前記検索結果のうち所定のデータに対して、順に、前記検索結果のデータが最上位になるように重みを付与し、各データに対する重みの

平均値を前記単語の重み値とすることを特徴とする付記9または10に記載の検索装置。

(付記14) 検索のための質問文を入力し、

検索対象のデータが格納されたデータベースを検索し、

前記入力された質問文に類似するデータを抽出し、

前記抽出された検索結果に対して、前記抽出に貢献した単語に関する貢献の度合を演算し、

前記演算された貢献度を対応する単語とともに出力することを特徴とする検索方法。(6)

(付記15) 前記出力は、前記貢献度に対応した表示形態で前記対応する単語を出力することを特徴とする付記14に記載の検索方法。

(付記16) 前記出力は、前記貢献度に対応したフォントで前記対応する単語を出力することを特徴とする付記14に記載の検索方法。

(付記17) 前記出力は、表示装置への出力であることを特徴とする付記15または16に記載の検索方法。

(付記18) 前記入力された質問文を単語に分割し、前記分割した単語毎に前記データベースを検索して類似するデータを抽出することを特徴とする付記14乃至17の何れか1項に記載の検索装置。

(付記19) 前記演算は、

前記抽出された検索結果のうち、前記類似の度合が上位である集団と前記類似の度合が下位である集団とを取得し、

前記分割した単語の前記上位集団に含まれる割合と前記下位集団に含まれる割合との差を求め、前記差を前記分割した単語の貢献度とすることを特徴とする付記18に記載の検索装置。

(付記20) 検索により質問単語に類似するデータを抽出し、前記抽出された検索結果に対して、前記抽出に貢献した単語に関する貢献の度合(貢献度)を対応する単語とともに画面上に表示する検索方法において、

前記画面上に表示された単語を指定し、

前記指定された単語に対して、前記検索のための重みを付与することを特徴と

する検索方法。

(付記 2 1) 検索のための質問文を入力し、

前記入力された質問文を単語に分割し、

前記分割した単語毎に検索対象のデータが格納されたデータベースを検索し、

前記入力された質問文に類似するデータを抽出し、

前記抽出された検索結果に対して、前記抽出に貢献した単語に関する貢献の度を演算し、

前記演算された貢献度を対応する単語とともに画面上に表示し、

前記画面上に表示された単語を指定し、

前記指定された単語に対して、前記検索のための重みを付与することを特徴とする検索方法。

(付記 2 2) 前記重み付けは、前記検索結果のデータが最上位になるように重みを付与することを特徴とする付記 2 0 または 2 1 に記載の検索方法。

(付記 2 3) 検索対象のデータが格納されたデータベースを検索し、

前記重みが付与された単語に類似するデータを抽出することを特徴とする付記 2 0 乃至 2 2 の何れか 1 項に記載の検索方法。

(付記 2 4) 検索対象のデータが格納されたデータベースを検索し、

前記重みが付与された単語に類似するデータを抽出し、

前記重み付けは、前記検索結果のうち所定のデータに対して、順に、前記検索結果のデータが最上位になるように重みを付与し、各データに対する重みの平均値を前記単語の重み値とすることを特徴とする付記 2 0 または 2 1 に記載の検索方法。

(付記 2 5) コンピュータに、

検索のための質問文を入力する機能、

検索対象のデータが格納されたデータベースを検索する機能、

前記入力された質問文に類似するデータを抽出する機能、

前記抽出された検索結果に対して、前記抽出に貢献した単語に関する貢献の度を演算する機能、

前記演算された貢献度を対応する単語とともに出力する機能、

を実現させるための検索プログラム。

(付記 2 6) 検索により質問単語に類似するデータを抽出し、前記抽出された検索結果に対して、前記抽出に貢献した単語に関する貢献の度合(貢献度)を対応する単語とともに画面上に表示するコンピュータに、

前記画面上に表示された単語を指定する機能、

前記指定された単語に対して、前記検索のための重みを付与する機能、

を実現させるための検索プログラム。

(付記 2 7) コンピュータに、

検索のための質問文を入力する機能、

前記入力された質問文を単語に分割する機能、

前記分割した単語毎に検索対象のデータが格納されたデータベースを検索する機能、

前記入力された質問文に類似するデータを抽出する機能、

前記抽出された検索結果に対して、前記抽出に貢献した単語に関する貢献の度合を演算する機能、

前記演算された貢献度を対応する単語とともに画面上に表示する機能、

前記画面上に表示された単語を指定する機能、

前記指定された単語に対して、前記検索のための重みを付与する機能、

を実現させるための検索プログラム。

【 0 0 7 3 】

【発明の効果】

以上説明してきたように、本発明によれば、検索要求中の単語の検索結果に対する貢献度の直感的な表示、ユーザーによる貢献度の容易な修正を実現でき、検索動作を制御することが可能となる。

【図面の簡単な説明】

【図 1】

本発明を適用した検索装置の機能構成図である。

【図 2】

本発明の原理説明図である。

【図 3】

検索結果中の単語の貢献度を表示するまでの処理の流れである。

【図 4】

検索結果中の単語の貢献度をユーザーが変更し、その変更内容を反映するように単語に重みを付ける処理の流れを示すフローチャートである。

【図 5】

検索要求の例を示す図であり、検索したいものを規定した文書である。

【図 6】

貢献度の計算式とパラメーターの設定値を示す図である。

【図 7】

貢献度の計算結果を示す図である。

【図 8】

貢献度を 4 階調のフォント表示とした図である。

【図 9】

単語貢献度への変更要求の例を示す図である。

【図 10】

再計算対象となる類似検索で上位にきた文書中に各単語がどのように含まれ、どのような得点となっているかを示す図である。

【図 11】

再計算の処理を示す図である。

【図 12】

重みの補正後の検索結果に対する貢献度を示す図である。

【図 13】

重み変更後の表示結果を示す図である。

【図 14】

検索装置の構成図である。

【図 15】

本発明におけるプログラムのコンピュータへのローディングを説明する図である。

【符号の説明】

- 1 検索装置
- 2 質問文入力手段
- 3 データベース
- 4 検索実行手段
- 5 単語貢献度演算手段
- 6 単語貢献度出力手段
- 7 上位集団下位集団取得手段
- 8 貢献度計算手段
- 9 単語指定手段
- 10 重み付け手段
- 11 再検索実行手段
- 12 検索結果出力手段
- 13 出力手段
- 21 質問文入力部
- 22 検索実行部
- 23 単語貢献度計算部
- 24 単語貢献度表示部
- 25 単語重み修正要求入力部
- 26 修正単語重み計算部
- 140 バス
- 141 CPU
- 142 メモリ
- 143 入力装置
- 144 出力装置
- 145 外部記録装置
- 146 可搬記録媒体
- 147 媒体駆動装置
- 148 ネットワーク接続装置

- 1 5 0 コンピュータ
- 1 5 1 メモリ (RAM、ハードディスク)
- 1 5 2 可搬型記録媒体 (CD-ROM、FLOPPY)
- 1 5 3 プログラム (データ)

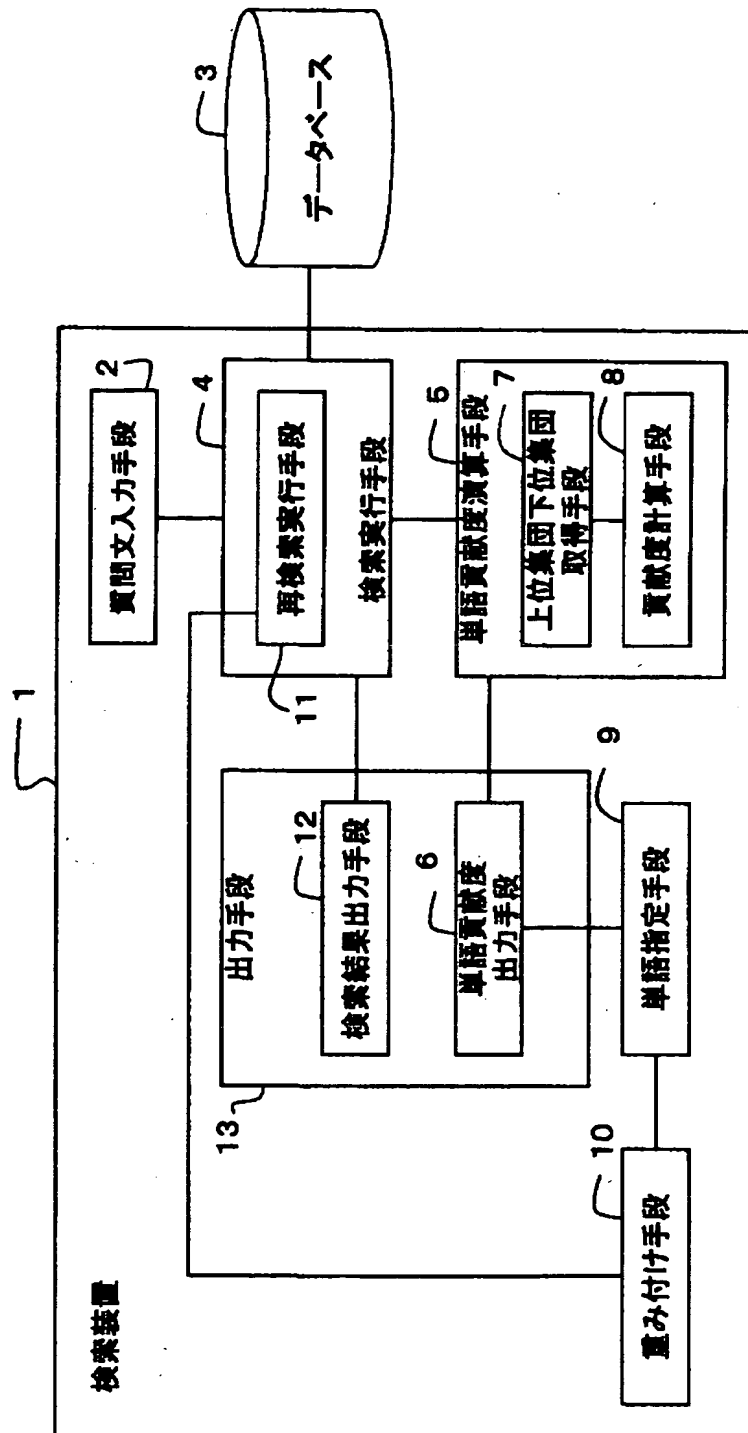
特 2 0 0 1 - 1 0 8 3 0 7

【書類名】

図面

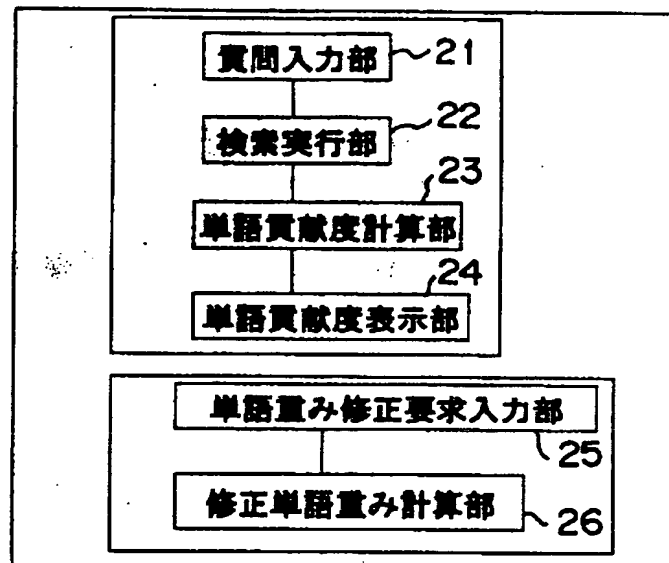
【図 1】

検索装置の機能構成図



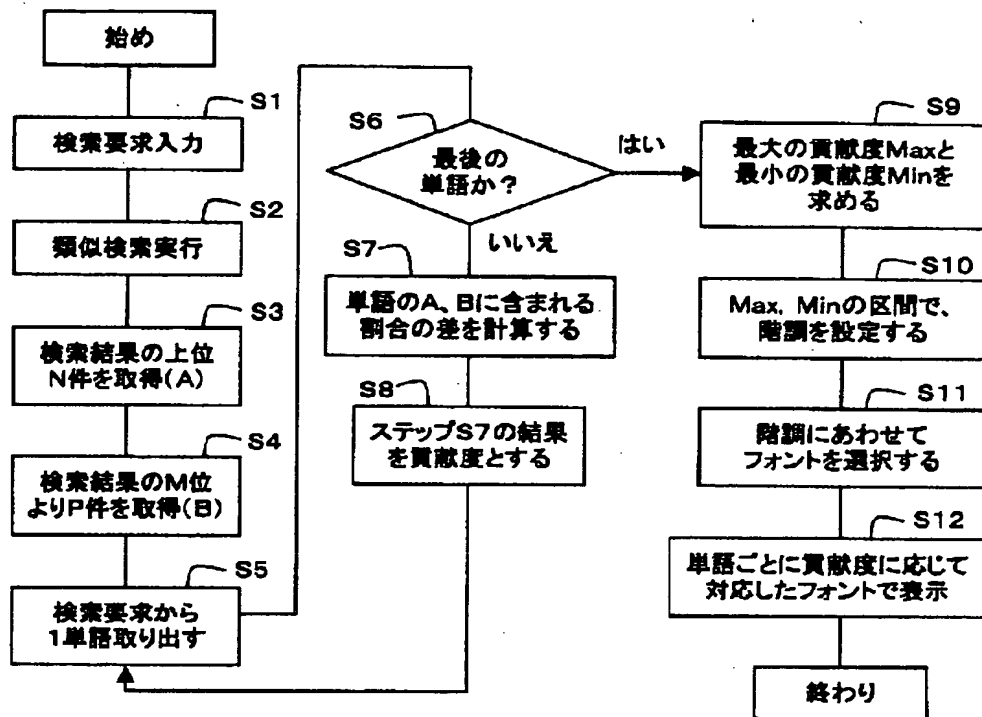
【図 2】

本発明の原理説明図



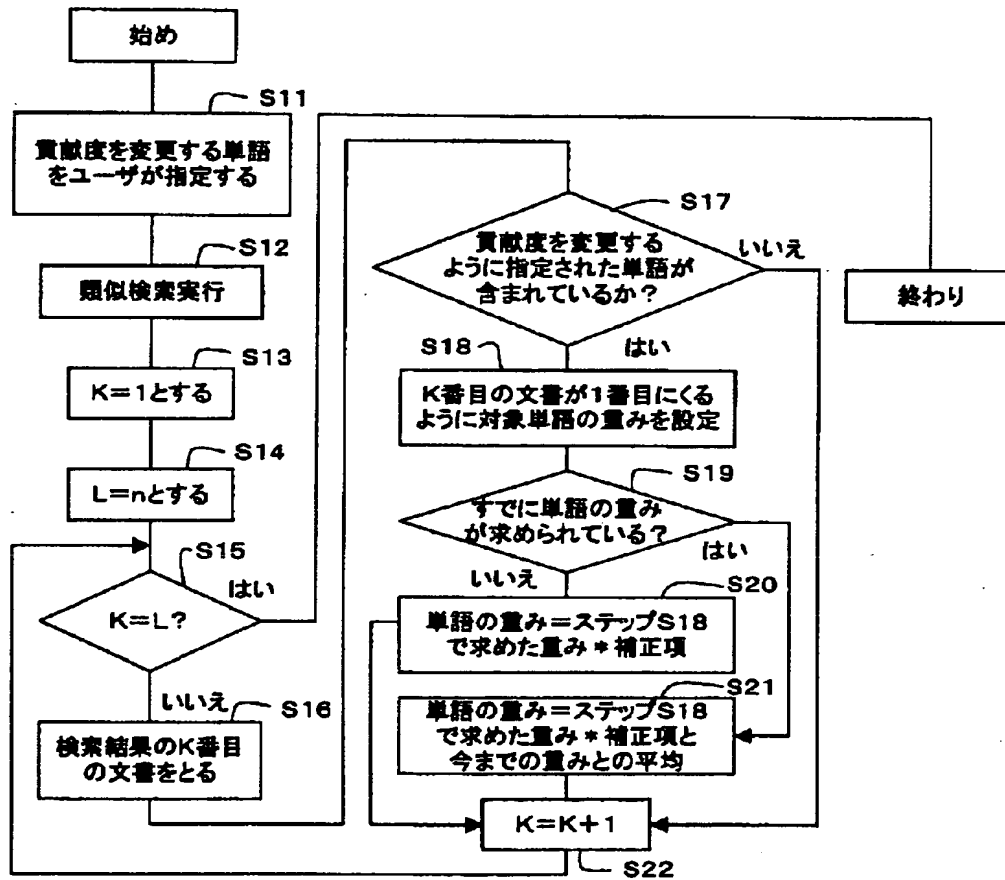
【図 3】

検索結果中の単語の貢献度を表示するまでの
処理の流れを示す図



【図 4】

検索結果中の単語の貢献度をユーザーが変更し、その変更内容を反映するように単語に重みを付ける処理の流れを示すフローチャート



【図 5】

検索要求の例を示す図であり、
検索したいものを規定した文書

- 目的

- hybrid car を検索する

- 質問文

- 自然言語で入力(サーチャーへの依頼形式)
 - TREC Query385番を例としてあげる
 - Identify documents that discuss the current status of hybrid automobile engines, (i.e., cars fueled by something other than gasoline only). A relevant document may include research on non-gasoline powered engines or prototypes that may be fueled by natural gas, methanol, alcohol; cost to the consumer; health benefits derived; and shortcomings in horsepower and passenger comfort.

【図 6】

貢献度の計算式とパラメーターの設定値を示す図

- A = 上位10件
- B = 上位800位から200件
- 計算式
- 貢献度計算式(Term Selection Value(Bougham formula))

$$Tsv = (r/R - \alpha * s/S) * w$$

$$\alpha = \text{parameter}$$

$$w = r/(R-r) / (n-r)/(N-n-R+r)$$

$$r = \text{Aの中で対象単語を含む文書数}$$

$$R = A$$

$$n = \text{対象単語を含む文書数}$$

$$S = B$$

$$s = \text{Bの中で対象単語を含む文書数}$$

$$N = \text{全文書数}$$

【図 7】

貢献度の計算結果を示す図

• fuel	7.2
• methanol	6.8
• cars	6.1
• gas	5.4
• automobile	5.1
• gasoline	4.8
• natural	4.5
• powered	3.7
• alcohol	2.4
• engines	2.2
• consumer	2.1
• passenger	2.0
• prototypes	1.6
• research	1.0
• benefits	1.0
• derived	0.9
• health	0.7
• hybrid	0.6

【図 8】

貢献度を4階調のフォント表示
とした図

1	Fuel, methanol, cars
2	gas, automobile, gasoline, natural
3	powered, alcohol, engines, consumer, passenger
4	prototype, research, benefits, derived, health, hybrid

【図 9】

単語貢献度への変更要求の例を
示す図

1	cars <u>hybrid</u>
2	gas, automobile, gasoline, natural
3	powered, alcohol, engines, consumer, passenger
4	Fuel, prototype, research, benefits, derived, health,

hybrid の貢献度を最大に上げる
fuel の貢献度を下げる
methanol を削除

【図10】

再計算対象となる類似検索で上位にきた
文書中に各単語がどのように含まれ、どの
ような得点となっているかを示す図

	Cars	hybrid	gas	automobile	gasoline	natural	powered	alcohol	fuel
1	5	0	1	1	3	1	2	1	5
2	4	0	1	0	3	1	2	1	5
3	3	0	0	0	2	1	1	0	3
:									
8	2	1	2	1	1	0	2	1	1
9									
10									
:									
:									
20									
:									
:									
:									
1000									

【図 11】

再計算の処理を示す図

	Cars	hybrid	gas	automobile	gasoline	natural	powered	alcohol	fuel
1	5	0	1	1	3	1	2	1	5
2	4	0	1	1	3	1	2	1	5
3	3	0	0	0	2	1	1	0	3
:									
8	2	1	2	1	1	0	2	1	1
9									

- Hybird を含む文書を上位に持ってくる計算
 - 1位の得点 19点, 8位の得点 11点
 - 8位を1位にするには, Hybrid に9を掛ければよい。
 - Hybrid の重みは $3 = 9 \cdot 1/\log(8)$
- fuel を含む文書を下位に持ってくる計算
 - 1位の得点 19点, 8位の得点 11点
 - 一位を8位に近づけるには, fuel を1(最低値に)
 - fuel の重みは $1/5 = 1/5 \cdot 1/\log(1)$

【図 12】

重みの補正後の検索結果に
対する貢献度を示す図

hybrid	5.3
cars	5.1
gas	4.8
automobile	4.6
gasoline	4.2
natural	4.1
fuel	4.1
powered	3.7
prototypes	3.4
alcohol	2.1
engines	1.3
consumer	1.1
passenger	1.0
research	0.9
benefits	0.5
derived	0.3
health	0.2

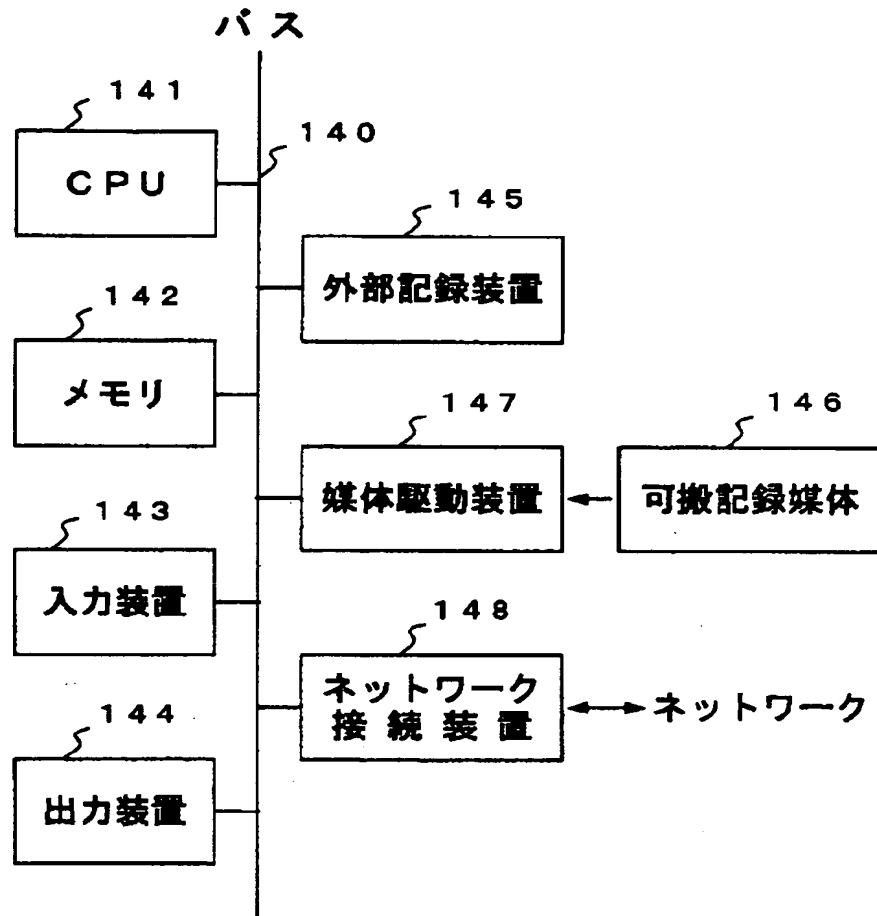
【図 13】

重み変更後の表示結果を示す図

• 1	hybrid ,cars ,gas , automobile
• 2	gasoline ,natural ,fuel ,powered,prototypes
• 3	alcohol engines consumer passenger
• 4	research , benefits, derived, health

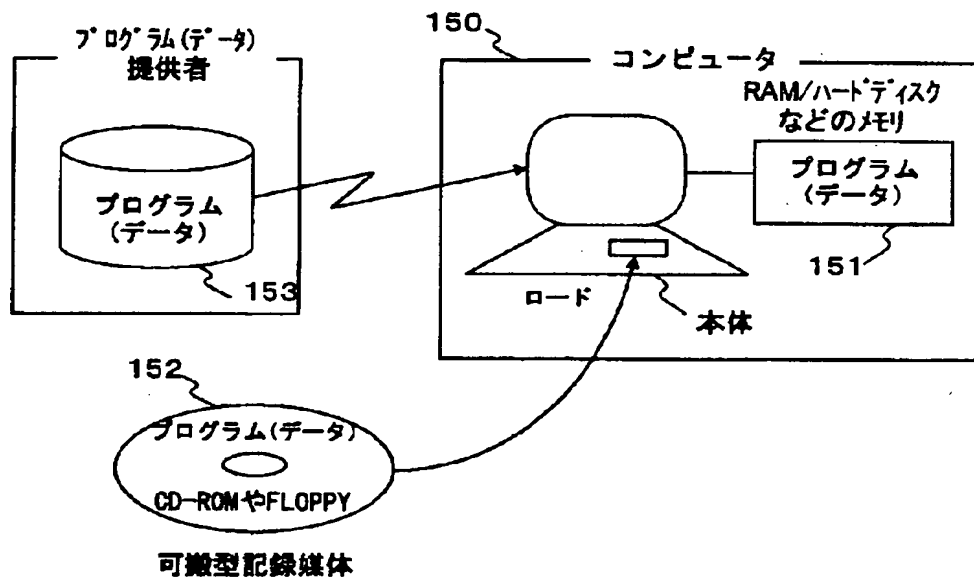
【図14】

検索装置の構成図



【図15】

プログラムのコンピュータへのローディングを説明する図



【書類名】 要約書

【要約】

【課題】 どのような単語が検索で重視されたかをユーザーに通知することができる検索装置、検索方法および検索プログラムを提供すること。

【解決手段】 検索のための質問文を入力する質問文入力手段と、検索対象のデータが格納されたデータベースを検索し、上記質問文入力手段によって入力された質問文に類似するデータを抽出する検索実行手段と、上記検索実行手段によって抽出された検索結果に対して、上記検索実行手段による抽出に貢献した単語に関する貢献の度合を演算する単語貢献度演算手段と、上記単語貢献度演算手段によって演算された貢献度を対応する単語とともに出力する単語貢献度出力手段とを備える。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000005223]

1. 変更年月日 1996年 3月26日

[変更理由] 住所変更

住 所 神奈川県川崎市中原区上小田中4丁目1番1号

氏 名 富士通株式会社